

# AGV通信プロトコルのQ-学習による獲得

旭川高専 ○加藤 龍 渡辺 美知子 古川 正志

## 要旨

AGVの自律運転には(1)経路の獲得、(2)通信プロトコルの獲得の2つの問題を解決する必要がある。本研究は、後者に焦点を当てAGVの通信プロトコル問題として、(1)速度の同調、(2)衝突回避を取り上げ、これらを複数AGVの協調問題としてQ-学習によるプロトコルの獲得方法を示す。また、数値実験からは貝原らのヒューリスティックな方法とほぼ同じ結果が得られることを示す。

## 1. はじめに

自律分散生産システムを構築する上での技術的課題として、各作業機械間を自由に移動できる自動搬送車(Automatically Guided Vehicles: AGVs)の自律的な行動をどの様に獲得するかが重要な問題となっている。この問題を解決することにより、従来の中央集権型の制御を不要とし、工場のスケジューリングをプログラムレス、かつ臨機応変に行うことが可能であると思われる。しかし、多数のAGVを同時にかつ自律的に動かすには、(1)経路の獲得(2)協調行動という大きな2つの問題が生じる。(1)はこれまでにQ-学習<sup>1)</sup>や確率的学習オートマトン等を用いた機械学習法が提案されている。(2)に関してはヒューリスティックな解法<sup>2)</sup>以外には今だ未解決である。本報告では、(2)を速度の協調、衝突回避としてのQ-学習を用いた通信プロトコル獲得問題としてとらえ、その獲得法を提案しその実験結果を示す。

## 2. Q-学習のアルゴリズム

Q-学習は状態と行動の対に対する評価の期待値(Q値)を用いて行動を選択し、その結果によりQ値を更新していく強化学習法である。ある時刻tの状態を $s_t$ 、その時の行動を $a_t$ とし、Q値を $Q_t(s_t, a_t)$ と置くと、行動決定はボルツマン分布により確率的になされる。

$$P(s_t, a_t) = \frac{\exp(Q_t(s_t, a_t)/T)}{\sum_{b \in A} \exp(Q_t(s_t, b)/T)} \quad (1)$$

ここでAは行動集合、Tは温度係数を表わす。Tが高くなるとQ値の差が反映されにくくなる。行動 $a_t$ を実行した結果、時刻t+1の時、状態 $s_{t+1}$ に移行したとすると $Q_{t+1}(s_t, a_t)$ は次式により更新される。

$$Q_{t+1}(s_t, a_t) = (1 - \alpha)Q_t(s_t, a_t) + \alpha \left[ c + \gamma \max_{b \in A} Q_t(s_{t+1}, b) \right] \quad (2)$$

ここで $\alpha$ は学習率で、これが高いほど学習速度が速くなる。 $\gamma$ は割引率で、これが高くなるほど長期的な報酬が考慮される。Cは環境からの評価を表わす。

## 3. エージェントのモデルによるAGVの定式化

AGVの自律走行をシミュレーションし、自律行動を獲得する為に、以下の様なエージェントモデルを作成した。

**3.1 エージェントの制約条件** 協調行動問題を考えるにあたって貝原らの方法同様、エージェント(AGV)が必要最小限のセンサと通信手段を用いて共同作業する場合

を想定し、かつ衝突問題のみを考えるため以下の条件を各AGVに与えることにする。

- a) AGVは次行程の作業装置の方角を知覚できる。
- b) AGVは自分の進行方向に隣接する他のAGVについてのみ、その進行方向のみ知覚できる。
- c) 作業機械は、その加工時間を0、AGVの収納数を $\infty$ とする。

**3.2 エージェントの内部モデル** AGV,  $A_i$  ( $i = 1, 2, \dots, TT$ ) を  $A_i = (P_i, P_m, D_i, ACT_i, FJ_i)$  と定義する。ここで、各項は以下に示される。

$P_i(x, y, t)$  : AGV $_i$ の時刻tにおける位置(x, y)

$P_m(x, y, t)$  : 作業機械 $_m$ の時刻tにおける位置(x, y)

$D_{it}$  : AGV $_i$ の時刻tにおける進行方向

{ $D_i$  | North, South, East, West}

$ACT_{it}$  : AGV $_i$ の時刻tにおける行動

{ $ACT_i$  | Forward, Right, Left}

$FJ_i$  : AGV $_i$ の作業完了機械数

$TT$  : 総AGV数

また貝原らの方法と比較するためにスケジューリングの価値を表わす評価値として、次式で示されるOperation Completion Ratio(OCR)を用いることにする。

$$OCR = \frac{\sum_{i=1}^{TT} FJ_i}{TT}$$

**3.3 AGV通信プロトコルの獲得** 協調速度の獲得は、局所的なAGVの周辺を観察し、もし他のAGVが自分の周りに存在するときにはそれを回避するか、速度を同調させるかの2つの行動が行われる。そこで、状態を相手の速度の状態  $D_{it} = D_{jt}$  (同方向),  $D_{it} \neq D_{jt}$  (異方向)の2種類とし、行動をACT {F, R, L}と設定し、Q値はこの状態-行動間において設定する。また、学習における環境からの評価は

if ( $P_i(x, y, t+1) \neq P_j(x, y, t+1) \cap P_i(x, y, t+1) \neq \forall P(x, y, t)$ )

$C_i$  (評価) = Reward (成功)

$C_j$  (評価) = Reward (成功)

if ( $P_i(x, y, t+1) = P_j(x, y, t+1) \cup P_i(x, y, t+1) = \forall P(x, y, t)$ )

$C_i$  (評価) = Penalty (失敗)

$C_j$  (評価) = Penalty (失敗)

else

$C_i$  (評価) = 0

すなわち、衝突回避には成功を、衝突には失敗を与えて、Q値を式(2)で更新する。

#### 4. シミュレーション実験

**4.1 仮想工場の設定** Fig.1のように100×100のフィールドを想定する。AGVおよび作業機械の初期位置は次式で与える。

$$P_i(x,y,0) = P_i(\text{random}(1\sim 100), \text{random}(1\sim 100), 0)$$

$$P_m(x,y,0) = P_m(\text{random}(1\sim 100), \text{random}(1\sim 100), 0)$$

但し、 $\text{random}(x_0\sim x_n)$  は $x_0\sim x_n$ からランダムに1つ選択する関数とする。

また、AGVが機械2でお互いに干渉させ衝突問題が生じるように、仕事順を機械(1→2→3)と機械(3→2→1)の2種類を与える。

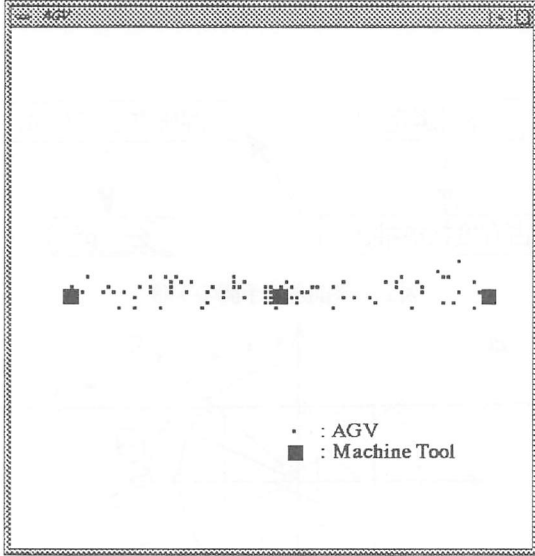


Fig.1 Experimental model

**4.2 数値実験** フィールドにAGVを100台発生させ、結果を見やすくするため、作業機械を一直線上に配置した。(Fig.1)  $\alpha$ を0.3,  $\gamma$ を0.8とし、これらの条件の下2つの実験を行った。1つ目の実験では、STEP数40000で行い、その学習状況をAGVの行動(Forward, Right, Left)の傾向として同方向、異方向の2種類について調べグラフ化した。その結果をFig.2, Fig.3に示す。Fig.2よりヒューリスティックな方法同様、同方向においてForwardの強い傾向が見られる。Fig.3では、異方向においてRight, Leftの相互関係を一定に保ち続けるような傾向が得られた。また、STEP数30000付近でRightに大きな乱れが生じているが徐々に元の安定な状態へ自律的に修復されている。2つ目の実験では1つ目の実験で得た通信プロトコルを用いた解法とヒューリスティックな解法とをOCRにより比較しグラフ化した。その結果をFig.4に示す。実験結果は貝原らのヒューリスティックな方法と比べほぼ同じ結果得られた。さらに衝突回避を無作為に行う方法と比較しても約20STEP近く速く収束することがわかった。

#### 5. 終わりに

本報告では自律分散システムの確立における問題解法のためのQ-学習を用いた通信プロトコルの獲得法について貝原らのヒューリスティックな方法と比較実験を行った。

報酬の与え方等のパラメータを慎重に選択する必要があったが、自律行動を可能とする結果が得られた。

今後は、本アルゴリズムを用いた高ロバスト性を有するシステムの構築し、実際の工場への実用化のため、変速による衝突回避などについてさらに検討する必要がある。

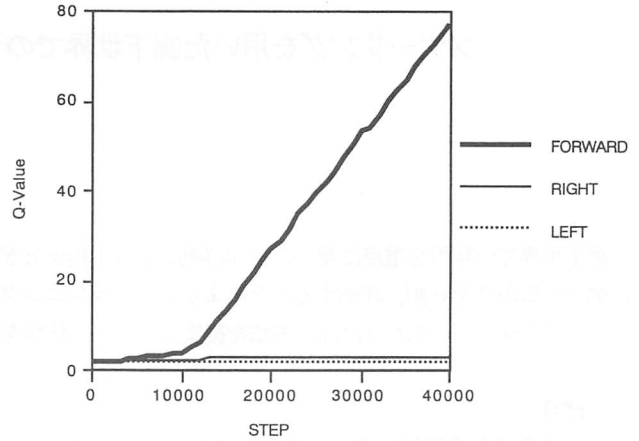


Fig.2 Two AGVs drive in the same direction.

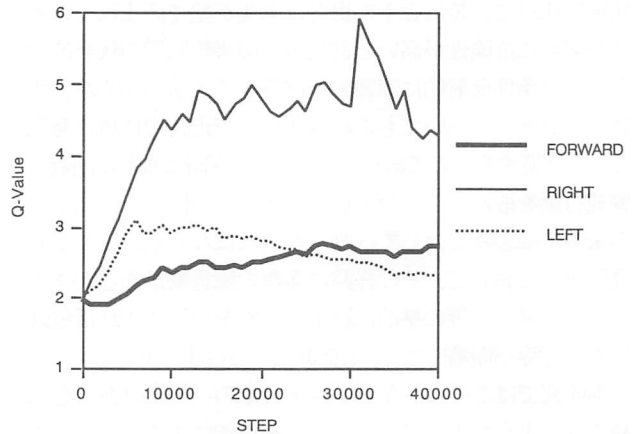


Fig.3 Two AGVs drive in the different direction, respectively.

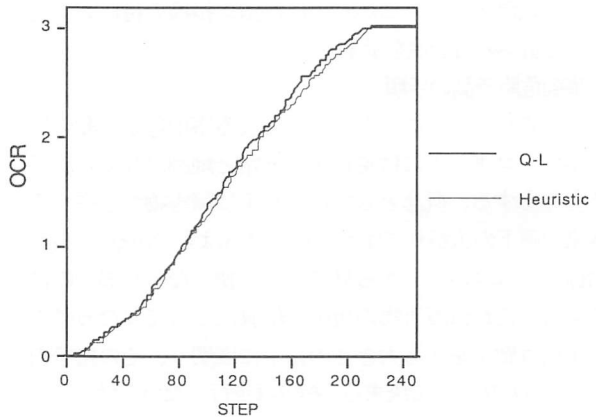


Fig.4 Comparison of Q-L result with heuristic by OCR

#### 参考文献

- 1) 谷 昌彦; 失敗と成功の競合におけるQ-学習の研究, 1997年度精密工学会北海道支部学術講演会講演論文集, (1997) 82-83
- 2) 貝原他; AN EVOLUTIONARY SELF-ORGANIZATION SCHEDULING USING COORDINATED AUTONOMOUS WORK AGENT, 1998 JAPAN-U.S.A. SYMPOSIUM ON FLEXIBLE AUTOMATION, vol.3 (1998) 1375-1381